植 物 分 类 与 资 源 学 报 2014, 36 (6): 707~714
Plant Diversity and Resources DOI: 10.7677/ynzwyj201414038

竹亚科 *lea3* 基因的进化分析*

于丽霞, 李 斌, 黄海泉, 鄢 波**
(西南林业大学园林学院, 云南 昆明 650224)

摘要: 选取竹亚科中两个超族、六个族和三个亚族的 10 个竹种为材料, 分别是泰竹、凤尾竹、青皮竹、大叶慈、慈竹、野龙竹、毛竹、香竹、苦竹、菲白竹, 分离克隆了它们的 *lea3* 基因, 并将它们与外类群物种水稻进行序列比对和进化分析。结果发现在分支模型与分支位点模型的检测中, 不同竹种所含 *lea3* 基因承受了不同的正选择压力, 清除选择作用在 *lea3* 基因编码区中占主导地位 ($\omega<1$)。在位点模型的检测中, 共检测出了 18 个显著性正选择位点, 占总氨基酸数目的 11.1%。对这 18 个显著性正选择位点进行定位后, 发现其中的 15 个位于 11 个氨基酸串联重复序列附近。这说明 *lea3* 基因中的 11 个氨基酸串联重复序列区比基因其它区域更容易受自然选择作用影响。同时, 在位点模型检测结果的基础上, 通过对强烈清除选择位点的定位, 发现在 11 个氨基酸串联重复序列区内存在一长段无强烈清除位点的序列区。

关键词: 竹亚科; *lea3* 基因; 基因克隆; 进化分析
中图分类号: Q 75 **文献标识码**: A **文章编号**: 2095-0845(2014)06-707-08

The Evolutionary Analysis of *lea3* Gene in Bambusoideae

YU Li-Xia, LI Bin, HUANG Hai-Quan, YAN Bo**
(*Southwest Forestry University*, *Faculty of Landscape Architecture*, Kunming 650224, China)

Abstract: The 10 bamboo species for cloning *lea3* gene concern two superfamily, six families and three subtribes, they are *Thyrsostachys siamensis*, *Bambusa multiplex*, *Bambusa textilis*, *Dendrocalamus farinosus*, *Neosinocalamus affinis*, *Dendrocalamus semiscandens*, *Phyllostachys heterocycla*, *Chimonocalamus delicates*, *Pleioblastus amarus*, *Sasa fortunei*. In the branch model and branch site model test, found that different bamboo species containing *lea3* gene under different selection pressure, purifying selection function is in dominant (omega<1). In the site model test, detected a total of 18 significant positive selection sites, account for 11.1% of the total amino acid number. Locating the 18 significant positive selection sites, found 15 sites located 11 amino acid tandem repeat sequences areas, this means that 11 amino acid tandem repeat sequences areas of *lea3* gene than in other regions more easily affected by natural selection effect. At the same time, based on the results of the site model test, through located these intence purifying selected sites, found a long no purifying selection sites area in 11 amino acid tandem repeat sequences.

Key words: Bambusoideae; *lea3* gene; Gene cloning; Evolutionary analysis

随着对 LEA 蛋白研究的深入, 通过分子杂交技术和免疫学方法, 可以将 LEA 蛋白分成 6 个族 (Battaglia 等, 2008; Dure 等, 1989; Dure, 1993; Ingram 和 Bartels, 1996)。第 3 族 LEA 蛋白含有多拷贝的由 11 个氨基酸组成的基元序列 (TAQAAKEKAGE) (Baker 等, 1988), 可形成 α-螺旋结构。在不同物种之间, 第三组 LEA 蛋白的同源性差异较大。例如, 大麦与小麦在氨基酸水平上同源性分别达 91% 和 95%, 但它们与同是禾本科的水稻、玉米、蒙古冰草的同源性却只

* 基金项目: 国家自然科学基金项目 (31160177, 31460202), 云南省部级重点学科、省高校重点实验室及校实验室共享平台资助
** 通讯作者: Author for correspondence; E-mail: yanbodr@yahoo.com.cn
收稿日期: 2014-03-13, 2014-06-13 接受发表
作者简介: 于丽霞 (1981-) 女, 硕士, 助理实验师, 主要从事植物分子生物学研究。E-mail: yulixia2005@sohu.com

有60%左右。第三组LEA蛋白在属内相对保守，而在属间变化较大（汤晓倩，2010）。

中国竹亚科共含有2超族、6族、5亚族、和43属；中国竹亚科分种（含变种、变型）检索表中，有708种、52变种、98变型和4个杂交竹种（易同培等，2009）。本研究中所用到的10个竹种分别属于竹亚科中的2个超族、6个族和3个亚族，它们分别是簕竹超族泰竹属的泰竹（*Thyrsostachys siamensis*）；簕竹超族簕竹族簕竹属的凤尾竹（*Bambusa multiplex*）与青皮竹（*Bambusa textilis*）；簕竹超族牡竹族慈竹属的慈竹（*Neosinocalamus affinis*）和牡竹属的大叶慈（*Dendrocalamus farinosus*），簕竹超族牡竹族牡竹属的野龙竹（*Dendrocalamus semiscandens*），簕竹超族倭竹族倭竹亚族刚竹属的毛竹（*Phyllostachys heterocycla*），北美箭竹超族丘斯夸竹族香竹属的香竹（*Chimonocalamus delicates*），北美箭竹超族北美箭竹亚族苦竹属的苦竹（*Pleioblastus amarus*），以及北美箭竹超族赤竹亚族赤竹属的菲白竹（*Sasa fortunei*）。

基因水平的适应性进化指一个遗传群体以一种具有较高适合度的等位基因替代另一种等位基因的过程。检测适应性进化有助于理解生物进化历史及相关生物结构与功能变异。在自然选择作用下，基因的适应性进化方式通常有三种，包括正选择、负选择和中性选择（徐鹏，2009）。在分子进化研究中，有效检测达尔文正选择的统计学方法是比较核苷酸序列上密码子的非同义替换（nonsynonymous substitution，氨基酸变化）与同义替换（synonymous substitution，氨基酸不变）比率（ω=dN/dS）。用比率可以检测编码蛋白序列所受到的选择压力，进而揭示三种不同的进化选择情况：当dN<dS（ω<1）时为负选择；当dN>dS（ω>1）时为正选择；当dN=dS（ω=1）时为中性突变（章张，2007）。

1 材料与方法

1.1 材料

1.1.1 *植物材料* 实验所用材料为幼嫩叶片，泰竹、凤尾竹、青皮竹、大叶慈、慈竹、野龙竹、毛竹、香竹、苦竹、菲白竹10个实验竹种均采于西南林业大学竹园，由该校孙茂盛老师鉴定。

1.1.2 *主要试剂与菌种* TIANGEN公司的植物总RNA提取试剂盒、TIANGEN公司的PCR产物回收纯化试剂盒、TAKARA公司的DL2000Marker；逆转录酶M-MLV（RNase H）RTase、oligo（dT）、dNTPs、Taq DNA polymerase、TAKARA公司的克隆载体pMD18-T Vector；PCR时所用到的各种试剂均购自于北京鼎国生物公司与TIANGEN公司，其它常规化学试剂购于北京鼎国生物公司。

大肠杆菌DH5α为本实验室保存菌种。

1.2 方法

1.2.1 *植物材料的处理* 将所摘取的新鲜竹叶枝条放置在室温下脱水2~4 h。

1.2.2 *RT-PCR和序列拼接* 对处理的竹子叶片采用液氮研磨法，利用RNAsimple Total RNA kit（北京天根）提取总RNA，并用First-Strand RT-PCR kit（大连宝生物）合成cDNA第一链，逆转录引物为dTP，其他按操作说明进行。

以上述合成的cDNA第一链为模板，应用禾本科植物*lea3*基因保守区域通用引物FP、RP扩增目的序列（Yu等，2010），扩增产物连接到pMD18-T载体上并进行测序，根据测序结果设计出第一轮3′ race引物F1（上游）：5′-AARGCGYRSGAGGCCAAGGAC-3′，3′ race引物R1（下游）：5′-CCAGTGAGCAGAGTGACGAGG-3′；第二轮3′ race引物F2（上游）：5′-GCCGGCAAGGACAAAACCGGC-3′，3′ race引物R2（下游）：5′-GACGAGGACTCGAGCTCAAGC-3′。利用套式PCR对3′端序列进行扩增，反应体系参照相关文献（于丽霞等，2010），两轮反应的退火温度均为62 ℃，模板为稀释10倍的第一轮PCR产物。所获得的3′端片段经克隆测序后，与已获得的相对保守片段拼接，得出部分cDNA序列。

1.2.3 *基因树的构建与分析* 用于建树的外类群为水稻。在建树之前，先对序列进行适当处理，具体方法如下：1）将核苷酸序列经MEGA中的“Translated Protein Sequences”转换成氨基酸序列；2）用其中的“Align by ClustalW”进行序列比对；3）将比对结果转换成“DNA Sequences”；4）手动去除其中的“Gap”；5）采用最大似然法（maximum likelihood，简称ML）构建。在构建树之前，bootstrap值全部设成1 000。然后将ML树文件及原有序列文件转入PAML软件（Yang等，2007）中。在调试好各种参数后运行“codeml.exe”控制文件。

1.2.4 *适应性进化检测* 分别采用了PAML4.1（Nei等，2000）中的3个模型：分支模型（Yang，1998）、位点模型（Nielsen和Yang，1998；Yang，2000），及分支位点模型（Yang和Wong，2005；Zhang等，2005）进行检测分析。

在分支模型中，进行了单比率（one-ratio）和自由比率（free-ratio）模型分析。单比率模型假定进化树上

所有分支的比率相同，属于最简单模型，用 ω_0 表示；自由比率模型是全类型模型，在该模型下假定不同分支有不同比率。该模型的优势在于，当缺少序列信息时，估测 ω 值不受影响，但由于参数较多，可能导致不精确的 ω 值出现。

在位点模型中，假定不同位点经历了不同选择压力，而在系统树上的不同分支上则无差异。在本研究中将 3 对模型的数据分别进行了似然比值检验（likelihood ratio test，LTR）。这三对模型为 M_1（近中性模型）vs M_2（正选择模型）、M_7（beta 模型）vs M_8（beta & ω 模型）、M_{8a}（beta）vs M_8（beta & v）（Nielsen 和 Yang，1998；Yang，2000）。在对三对模型的数据进行 LRT 检验时，使用统计量 $-2\Delta\ell$（其中的 $\Delta\ell=\ell_1-\ell_0$，ℓ_1 和 ℓ_0 分别表示两个模型下的似然值）依 χ^2 卡方分布检验显著性，自由度为两个模型参数的数值之差。

在分支位点模型中，根 d_N/d_S 的四类比率区分不同位点。第一类由所有分支中高度保守的位点组成，其中的 d_N/d_S 比率 ω_0 较小；第二类由处于中性进化的位点组成，其 d_N/d_S 比率 ω_1 等于 1；第三类由背景分支高度保守（$0<\omega_0<1$），同时前景分支比率 ω_2 显著大于 1 的位点组成；第四类指背景分支处于中性进化（$\omega_0=1$），但前景分支比率 ω_2 显著大于 1 的位点组成。其中第三类与第四类中的 ω_2 数值相同。分支位点模型包括 A 和 B 两个模型（Nielsen 和 Yang，1998；Yang，2000），A 模型下又包括 2 个检验：检验 1 和检验 2。通常采用检验 2。在检验 2 中，零假设下假定所有分支中只有两类位点，$0<\omega_0<1$ 和 $\omega_0=1$。控制文件中需要设定 fix_omega = 1 及 omega = 1。备选假设比零假设多 2 个参数 p_2 和 ω_2，其中的 ω_2 大于或等于 1，如结果 $\omega_2>1$ 就说明前分上存在正选择位点。

2 结果与分析

2.1 竹亚科 *lea3* 基因拼接序列的获得

利用通用引物 FP 和 RP 经 RT-PCR 成功地从竹子的 cDNA 中扩增出长度约为 500 bp 左右的片段，与预计扩增的长度相符。利用已获得的这段序列设计了 3′RACE 引物，第一轮 3′-Race-PCR 扩增的产物经琼脂糖凝胶电泳检测，结果凝胶电泳图上的泳道中均显示出明亮的“Smear”状条带，这说明了在 RACE-PCR 条件下用 3′端引物扩展特异性差，目的条带不明显。以第一轮扩增产物稀释 10 倍为模板，进行第二轮 3′-Race-PCR，将扩增的产物分别回收、纯化、及克隆测序，与通用引物扩增的保守区域片段拼接，得到了部分 cDNA 序列，大小分别为泰竹 540 bp、凤尾竹 531 bp、青皮竹 540 bp、大叶慈竹 552 bp、慈竹 552 bp、野龙竹 552 bp、毛竹 585 bp、香竹 552 bp、苦竹 552 bp、菲白竹 546 bp。

2.2 基因树的构建

最大似然法是根据特定替代模型分析既定的一组序列数据，从而使所获得的每一个拓扑结构的似然值最大，然后挑选出其中似然值最大的拓扑结构作为最优系统树（Huelsenbeck 和 Rannala，1997；Yang，1998）。根据已有资料显示，如果模型合适，ML 效果较好（Saiton 和 Nei，1987；Hasegawa 等，1991）。将用于比对的序列经 MEGA 中的“Translated Protein Sequences”转换后，用其中的“Align by ClustalW“进行序列比对，手动去除“Gap”，最后将这些序列用于建树。以 ML 法构建的进化树为树文件，利用 PAML 软件对 11 条 *lea3* 基因序列进行进化分析（图 1）。从图中可以看出，禾本科稻属的水稻与任何一种竹子的亲缘关系均较远，慈竹和大叶慈亲缘关系较近，这与它们同属牡竹族相吻合；而同属簕竹属的凤尾竹和青皮竹亲缘关系却并不很近，这可能与它们的生境不同有关

2.3 竹亚科 *lea3* 基因的选择压力检测结果

2.3.1 正选择压力检测结果　单比率模型（oneration）假定所有分支都具有相同的 ω 值，涉及参数 21 个。在该模型下，似然值为 $\ell_0=-2296.944859$，$K=1.63425$，估测 ω 值为 0.36203（表 1）。在分支模型中，自由比率模型（free-ratio）假定每个分支都有独立的 ω 值，包括参数 39 个，似然值为 $\ell1=-2289.125585$。因为自由比率模型比单比率模型多 18 个参数，所以 $2\Delta\ell=2(\ell_1-\ell_0)=15.65$，自由度 $df=18$。同时，数据显示不同分支所受选择压力各异，多数分支处于强烈清除选择压力下（$\omega<1$），少数分支处于正选择压力下（$\omega>1$）。通过似然比值检验分析显示（表 2），在自由比率模型 vs. 单比率模型的检测中，$2\Delta\ell=15.64$，$P=0.618$，$df=18$。

在分支位点模型中，a-s 表示分支，当 p^* 或 r^* 表示前景分支时，其它分支为背景分支。在此模型下，分支 r 检测到两个正选择位点（$P>0.9$：34K $P>0.6$：32 D）（表 1）。通过似然比值检验分析显示（表 2），在以 r 为前景分支的零假设 vs. 备选假设检测中，$2\Delta\ell=3.82$，$P=0.051\approx5\%$，$df=1$。

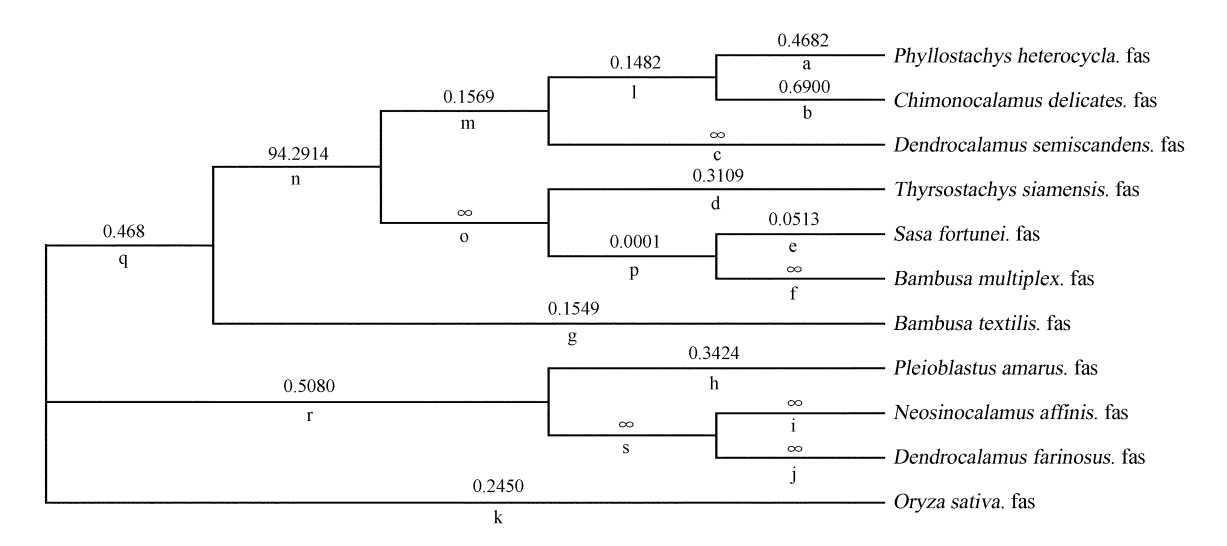


图1 用PAML分析的ML树。1 000次重复的自展支持率，分支上的数字表示自由比率模型“free-ratio”得到的 ω 值，a-s为分支代号

Fig. 1 ML tree of the PAML analysis. 1 000 replicates bootstrap support rate. Branch on the numbers indicate the free-ratio model ω value in “free-ratio”, a-s: the branch code

表1 竹亚科植物 *lea3* 基因片段在不同模型下的参数估计值以及对数似然值

Table 1 Parameter estimates and Log-Likelihood values of the *lea3* gene in gymnosperm under different models

模型 Model	参数个数 P	似然值 ℓ	ts/tv	参数估测值 Estimates of parameters	正选择位点 Positively selected sites
M0：单比率 (M0：one-ratio)	21	−2296.95	1.63	$\omega_0=0.36$	None
分支模型					
自由比率 F：(Free-ratio)	39	−2289.13	1.63	ωj=∞ ωi=∞ ωr=∞ ωp=∞ ωc=∞	None
位点模型					
M1：中性 (M1：neutral)	22	−2104.52	2.27	p0=0.60，(p1=0.40) ω0=0.03，ω1=1	Not allowed
M2：选择 (M2：selection)	24	−2063.59	2.43	p0=0.53，p1=0.29，p2= 0.18 ω0=0.04，ω1=1，ω2=8.32	$P>0.99$：24 R、29 T、47 K、54 A、58 K、59 A、62 A、69 R、71 H、76 Q、81 I、82 G、104 L；$P>0.95$：50 G、56 K、79 S
M7：(beta)	22	−2101.34	2.25	p=0.16，q=0.29	Not allowed
M8：(beta & ω)	24	−2061.34	2.42	p0=0.81，p=0.18，q= 0.45，(p1=0.19) ω=6.74	$P>0.99$：24 R、29 T、47 K、54 A、58 K、59 A、62 A、69 R、71 H、76 Q、81 I、82 G、104 L；$P>0.95$：50 G、56 K、57 E、60 F、79 S
M8a	23	−2096.98	2.28	p0=0.71，p=0.31，q=3.46，(p1= 0.29) ω=1	None
分支位点模型					
模型A (model A)					
q*	24	−2104.52	2.27	p0=0.60，p1=0.40，(p2+p3=0) ω2= 1	None
r*	24	−2101.81	2.29	p0=0.57，p1=0.39，(p2+p3=0.04) ω2=12.72	$P>0.9$：34K $P>0.6$：32 D

注：P 为模型中参数的数目；ω_0 是指所有分支上的 d_N/d_S，ω_q 与 ω_r 分别是分支q与s分支上的 d_N/d_S；q*与r*表示在用分支位点模型检测正选择位点时，分支q与分支r是前景分支

Note：P number of parameters in the model；ω_0 is the d_N/d_S ratio for all branches，ω_q and ω_r are the d_N/d_S ratios for branches q and r；branches q and r are the foreground branches

表 2　似然比值检验统计量（2Δℓ）

Table 2　Likelihood ratio statistics（2Δℓ）

模型对比 Comparison	2Δℓ	自由度 *df*	概率 *P*
F vs. M0 模型	15.64	18	0.618
M1（中性）vs. M2（选择）	81.85	2	1.68586E-18
M7（beta）vs. M8（beta & v）	80.01	2	4.23239E-18
M8a（beta）vs. M8（beta & v）	71.28	1	3.0992E-17
q*：零假设 vs. 备选假设	0	1	1
r*：零假设 vs. 备选假设	3.82	1	0.051

注：显著：$P<5\%$，$\chi^2=3.84$；极显著：$P<1\%$，$\chi^2=6.63$。E-18 = 10^{-18}

Note：Significant：（$P<5\%$，$\chi^2=3.84$）；Extremely significant：（$P<1\%$，$\chi^2=6.63$）. E-18 = 10^{-18}

在位点模型中，M2 模型和 M8 模型均检测到 13 个一样的极显著正选择位点（$\omega>1$，$P>0.99$，占 8.02%）（表 1），及多个显著正选择位点（$\omega>1$，$P>0.95$，占 3.09%）。这 18 个正选择位点的定位见图 3。同时，在 M8 模型下还检测到 62 个强烈清除选择位点，占总氨基酸位点数的 38.3%。经分析后发现，在这些强烈清除选择位点间存在一段长段无强烈清除选择位点的区域（图 2 中方框部分）。似然比值检验结果显示（表 2）：M1 vs. M2：$2\Delta\ell=81.85$，$P=1.68586\text{E-}18<1\%$，$df=2$；M7vs. M8：$2\Delta\ell=80.01$，$P=4.23239\text{E-}18<1\%$，$df=2$；M8a vs. M8：$2\Delta\ell=71.28$，$P=3.0992\text{E-}17<1\%$，$df=1$。

2.3.2　负选择压力检测结果　本研究中共检测到了 65 个后验平均 ω 值小于 0.1 的极强烈负选择位点，同时发现在 46D-74T 之间存在一长段没有极强烈负选择位点出现的区域（结果如表 3 所示）。

表 3　M_8模型下清除选择位点的后验平均 ω 值

Table 3　Posterior mean ω for the purifying selected sites under M_8 model

位点 Site	后验平均 ω 值 Posterior Mean ω	位点 Site	后验平均 ω 值 Posterior Mean ω	位点 Site	后验平均 ω 值 Posterior Mean ω
1 A	0.070	75 E	0.061	126 V	0.063
2 G	0.062	84 T	0.067	127 G	0.062
3 E	0.061	87 A	0.070	128 A	0.070
5 K	0.061	88 A	0.072	129 K	0.061
9 E	0.061	89 K	0.063	130 D	0.061
10 E	0.061	90 Q	0.081	131 A	0.071
11 K	0.061	91 K	0.061	132 V	0.063
13 G	0.063	96 A	0.071	133 M	0.069
17 G	0.062	98 Y	0.078	135 T	0.068
22 K	0.072	100 K	0.061	137 G	0.063
23 A	0.071	105 A	0.070	138 M	0.069
25 E	0.061	106 G	0.062	141 D	0.061
26 A	0.070	107 K	0.061	143 G	0.062
27 K	0.061	108 D	0.061	146 D	0.094
28 D	0.061	110 T	0.067	150 K	0.061
35 E	0.061	111 G	0.062	151 D	0.061
36 A	0.071	113 V	0.062	152 T	153 S
37 T	0.067	116 Q	0.081	154 A	0.070
38 K	0.061	117 A	0.070	155 T	0.078
41 A	0.070	119 E	0.061	156 E	0.061
46 D	0.061	120 Q	0.081	160 R	0.065
74 T	0.067	121 V	0.063		

1. *N. affinis*. fas	AGEAKARAEEKTGYVMGAAQEKAREAKDTAS-----------DAKEATKEKACE
2. *P. heterocycla*. fas	AGEAKGRTEEKVGHAMGVTKDKAYEAKDRGSVAAGHAAGEGHDAKEATKG----
3. *D. farinosus*. fas	AGETKGRAEEKTGYVMGAAQEKAREAKDTAS-----------DAKEATKEKACE
4. *S. fortunei*. fas	AGETKARTEEKAGHVMGSTKDKAYEAKDRASDAAGQAAGKGHDAKEATKDKAYE
5. *T. siamensis*. fas	AGETKGRTEEKAGHVMGATKDKAYEAKDRALDAAGQAAGKGHDAKEATKDKAYE
6. *B. multiplex*. fas	AGEAKGRTEEKAGHVMGSTKDKAYEAKDRASDAAGQAAGKGHDAKEATKDKAYE
7. *B. textilis*. fas	AGETKGRTEEKAGHVMGATKDKAYEAKDRASDAAGQAAGKGHDAKEATKDKAYE
8. *D. semiscandens*. fas	AGETKGRTEEKAGHVMGATKDKAYEAKDRASDAAGQAAGKGHDAKEATKDKAYE
9. *C. delicates*. fas	AGETKGRTEEKAGHAMGVTKDKAYEAKDRGSDAAGHAAGEGHAAGEATKGKAWE
10. *P. amarus*. fas	AGEAKGRAEEKAGYVIGAAREKAREAKDTASDAAGRAMGRGQDAKEATKEKAYE
11. *O. sativa*. fas	AGETKAHTEEKAGQVMGASKDKASEAKDRASEAAGHAAGKGQDTKEATKEKAQA

1. *N. affinis*. fas	AKDRAADATGRATDKGRGTADATKEKAFEAKDKAADTAQAARERAHDGTEQTGS
2. *P. heterocycla*. fas	---KASETAQAAKDRACEGKDQTGSFLGQK-------TEAAKQKASETTEATKQ
3. *D. farinosus*. fas	AKDRAADATGRATDKGRGAADATKEKAFEAKDKAADTAQAARERAHDGTEQTGS
4. *S. fortunei*. fas	AKDKASETAQAAKDRAGEGKDQTGSFLGQK-------TEAAKQKASETTEATKQ
5. *T. siamensis*. fas	AKDRASETAQAAKDRAGEGKDQTGSFLGQT-------TEAAKQKASETTEATKQ
6. *B. multiplex*. fas	AKDKASETAQAAKDRAGEGKDQTGSFLGQK-------TEAAKQKASETTEATKQ
7. *B. textilis*. fas	AKDKASETAQAAKDRAGGGKDQTGSFLGQK-------TEAAKQKASETTEATKQ
8. *D. semiscandens*. fas	AKDKASETAQAAKDRAGGGKDQTGSFLGQK-------TEAVKQKASETTEATKQ
9. *C. delicates*. fas	AKDKASETAQAAKDRACEGKDQTGSFLGQK-------TEAAKQKASETTEATKQ
10. *P. amarus*. fas	AKDRAADAT----DKGRGAADATKEKACEA-----DTAQAARDRARDGTEQTES
11. *O. sativa*. fas	AKERASETAQAAKDKTSSTSQAARDKAAES-------KDQTGGFLGEKTEQAKQ

1. *N. affinis*. fas	YIGQT-----------ADAAKQKAADAAQYAKETALAGKDKTGSVLQQAGEQVK
2. *P. heterocycla*. fas	KASET-----------AEAAKQKTSDAAQYTKDSAVAGKDKTGAVLQQAGEQVK
3. *D. farinosus*. fas	YIGQT-----------ADAAKQKAADAAQYAKETAPAGKDKTGSVLQQAGEQVK
4. *S. fortunei*. fas	KAAET-----------TEAAKQKTSDAAQYAKDSAVAGKDKTGSVLQQAGEQVK
5. *T. siamensis*. fas	KAAET-----------TEAAKQKTSDAAQYAKDSAVAGKDKTGSVLQQAGEQVK
6. *B. multiplex*. fas	KAAET-----------TEAAKQKTSDAAQYAKDSAVAGKDKTGSVLQQAGEQVK
7. *B. textilis*. fas	KAAETTEATKQKAAETTEAAKQKTSDAAQYAKDSAVAGKDKTGSVLQQAGEQVK
8. *D. semiscandens*. fas	KASET-----------TEAAKQKTSDAAQYAKDSAVAGKDKTGSVLQQAGEQVK
9. *C. delicates*. fas	KASET-----------AEAAKQKTSDAAQYTKDSTVAGKDKTGAVLQQAGEQVT
10. *P. amarus*. fas	YVGQT-----------AEAAKQKAAGAALYAKETALAGKDKTGAVLQQAGEQVK
11. *O. sativa*. fas	KAAET-----------AGAAKQKTAETAQYTKDSAIAGKDKTGSVLQQASEQVK

1. *N. affinis*. fas	NVAVGAKDAVMNTLGMAGDK-GSDDSTTKDTS----ATETITRDR
2. *P. heterocycla*. fas	SAAVGAKDAVMNTLGMGGDK-GGADSTTKDTS----ATETITRDR
3. *D. farinosus*. fas	NVAVGAKDAVMNTLGMAGDK-GSDDSTTKDTS----ATETITRDR
4. *S. fortunei*. fas	NVAVGAKDAVMNTLGMAGDK-GSDDSTTKDTS----ATETITRDR
5. *T. siamensis*. fas	NVAVGAKDAVMNTLGMAGDK-GSDDSTTKDTS----ATETITRDR
6. *B. multiplex*. fas	NAAVGAKDAVMNTLGMAGDK-GSDDSTTKDTS----ATETITRDR
7. *B. textilis*. fas	NAAVGAKDAVMNTLGMAGDK-GSDDSTTKDTS----ATETITRDR
8. *D. semiscandens*. fas	NVAVGAKDAVMNTLGMAGDK-GSDDSTTKDTS----ATETITRDR
9. *C. delicates*. fas	SAAVGAKDAVMNTLGMGGDK-GRADSTTKDTS----ATETITRDR
10. *P. amarus*. fas	SAAVGAKDAVMNTLGMGGDK-GGADSTTKDTS----ATETITRDR
11. *O. sativa*. fas	STVVGAKDAVMSTLGMTEDEAGTDDGANKDTSATAAATETTARDH

图2 竹亚科与外类群水稻的氨基酸序列比对图

注：方框内的为无强烈清除选择位点区域；“＊”号为正选择位点；底部的粗下划线部分代表含有11个氨基酸的串联重复区；双下划线代表含有11个氨基酸的串联重复基元序列；单下划线代表含有相对稳定结构的特殊序列；三下划线代表有疑问的11个氨基酸的串联重复基元序列

Fig. 2 Bamboosubfamilies and outgroup *Oryza sativa*'s acid sequence comparing figure

Note：inside the box for strong clear select site region；“＊”for positive selection sites；underline the 11 amino acids by tandem repeats；double underline representation containing 11 amino acids by tandem repeat sequence of primitive elements；single underline representative contain relatively stable structure of insertion sequences；three underline on behalf of questionable 11 amino acids of the tandem repeat sequence of primitive elements

2.4　竹亚科 *lea3* 基因的检测结果定位

将检测到的正选择位点、无强烈负选择位点的长段区域、11 个氨基酸的串联重复序列及特殊序列在氨基酸序列比对图上进行定位。定位后的结果如图 2 所示。图中 11 个氨基酸串联重复序列区包括 91 个氨基酸，占总长度 207 个氨基酸的 44%左右（图中加粗线的部分）。

本研究检测出的 18 个正选择位点中 15 个位于 11 个氨基酸的串联重复序列区（下图中加粗线的部分），3 个位于插入片段区（下图中加单下划线的部分）。检测到的无强烈负选择位点的长段区域共含有 34 个氨基酸位点，位于 11 个氨基酸串联重复序列中间位置（下图中用方框标记的部分）。

3　讨论

3.1　竹亚科 *lea3* 基因中检测到的正选择分支

分支模型的检测结果显示，不同竹种所含 *lea3* 基因所承受的正选择压力不同。原因可能是由于不同竹种所处的生境不同，从而经历了不同的自然选择作用。在分支位点模型中，将分支 q 与分支 r 进行配对检测，在分支 q 上没有检测到正选择位点，在 r 分支上检测到两个正选择位点。这说明不同竹种因所处生境的不同而经历了不同的自然选择作用。

3.2　竹亚科 *lea3* 基因中检测到的正选择位点

在位点模型中共检测到 5 个显著正选择位点和 13 个极显著正选择位点。相比于其它基因，*lea3* 基因中所存在的正选择位点数目较多。已有研究结果表明，抗病及抗旱类奢侈基因（luxury gene）中所检测到的正选择位点数明显多于与光合作用有关的看家基因（house-keeping gene）（罗莎，2012）。这显示抗病及抗旱类奢侈基因相比于与光合作用有关的看家基因更容易受到自然选择作用影响。根据中性进化原理，基因突变率在没有受到强烈理化因子刺激时几乎不变。针对竹亚科 *lea3* 基因中存在较多正选择位点的现象，可能是由于该基因属于奢侈基因，相比于其它看家基因所受到的自然选择压力较小，位点突变更容易被保留下来。

通过对 18 个正选择位点进行定位，发现其中 15 个极显著正位点位于 11 个氨基酸串联重复序列区内，3 个位点位于 7 个特殊的氨基酸序列区。这说明 11 个氨基酸串联重复序列区比基因其它部分更容易受自然选择作用影响。原因推测可能与 11 个氨基酸串联重复序列所形成的特殊蛋白质结构有关。11 个氨基酸串联重复序列可以形成 α-螺旋结构，而 α-螺旋的疏水界面又可形成同型二聚体结构，这种同型二聚体结构能在植物受到干旱胁迫时保护生物大分子，减轻水份胁迫伤害，与植物抗旱性密切相关。

3.3　竹亚科 *lea3* 基因中检测到的强烈清除选择位点

将负选择位点数目与正选择位点数目进行比较，发现前者远大于后者。这说明 *lea3* 基因尽管受到较强正选择压力，但清除选择作用依然占主导地位。对所检测到的 65 个强烈清除选择位点进行观察后，发现其间存在一长段无强烈清除选择位点的区域。这段无强烈清除选择位点的区域位于靠近 11 个氨基酸串联重复序列区中心部位，同时，该区域也是正选择位点最集中的区域。这说明这段无强烈清除选择位点的区域所受到的清除选择压力极小，正选择压力占主导地位。

〔参　考　文　献〕

易同培，马丽莎，史军义等，2009. 中国竹亚科属种检索表［M］. 北京：科学出版社

章张，2007. 基于进化选择核动力的算法研究［D］. 北京：中国科学院计算技术研究所

徐鹏，2009. 社会性昆虫抗菌肽基因的分子进化研究［D］. 杭州：浙江大学

Baker J，Steele C，Dure LIII，1988. Sequence and characterization of 6 lea proteins and their genes from cotton［J］. *Plant Molecular Biology*，**11**：277—291

Battaglia M，Olvera-Carrillo Y，Garciarrubio A *et al.*，2008. The enigmatic lea proteins and other hydrophilins［J］. *Plant Physiology*，**14**（1）：6—24

Dure L，Crouch M，Harada J *et al.*，1989. Common amino acid sequenc domains among the lea proteins of higher plants［J］. *Plant Molecular Biology*，**12**（5）：475—486

Dure L，1993. A repeating 11-mer amino acid motif and plant desiccation［J］. *The Plant Journal*，**3**（3）：363—369

Hasegawa M，Kishino H，Saitou N，1991. On the maximum likelihood method in molecular phylogenetics［J］. *Journal of Molecular Evolution*，**32**（5）：443—445

Huelsenbeck JP，Rannala B，1997. Phylogeny methods come of age：

testing hypotheses in an evolutionary context [J]. *Science*, **276** (5310): 227—232

Ingram J, Bartels D, 1996. The molecular basis of dehydration tolerance in plants [J]. *Annual Review of Plant Physiology and Plant Molecular Biology*, **47**: 377—403

Lu S (罗莎), 2012. Evolutionary analysis of the *NBS-LRR* encoding grnes in Poaseae (PhD) [D]. Wuhan: Huazhong Agricultural University

Nielsen R, Yang Z, 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene [J]. *Genetics*, **148**: 929—936

Saitou N, Nei M, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees [J]. *Molecular Biology and Evolution*, **4** (4): 406—425

Tang XQ (汤晓倩), Yu LX (于丽霞), Wu XL (武晓璐) *et al*., 2010. Rseearch advance in the group three of late-embryogenesis-abundant proteins and genes [J]. *Chinese Bulletin of Life Science* (生命科学), **22** (6): 551—555

Yang X, Stojkovic EA, Kuk J *et al*., 2007. Crystal structure of the chromophore binding domain of an unusual bacteriophytochrome, RpBphP3, reveals residues that modulate photoconversion [J]. *Proceedings of the National Academy of Sciences of the United States of America*, **104** (30): 12571—12576

Yang Z, 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution [J]. *Molecular Biology and Evolution*, **15** (5): 568—573

Yang Z, 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A [J]. *Journal of Molecular Evolution*, **51**: 423—432

Yang Z, Wong WSW, 2005. Nielsen R, Bayes empirical Bayes inference of amino acid sites under positive selection [J]. *Molecular Biology and Evolution*, **22**: 1107—1118

Yu L, Wu X, Tang X *et al*., 2010. Simple and efficient method for isolating cDNA fragments of *lea3* genes with potential for wide application in the grasses (Poaceae) [J]. *Genetics and Molecular Research*, **9** (3): 1321—1325

Yu LX (于丽霞), Wu XL (武晓璐), Tang XQ (汤晓倩) *et al*., 2010. Effect of new specificity and yield enhancer on amplification of *lea3* gene [J]. *Acta Botanica Yunnanica* (云南植物研究), **32** (6): 535—538

Zhang J, Nielsen R, Yang Z, 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level [J]. *Molecular Biology and Evolution*, **22** (12): 2472—247